Embedded Video Coding Using Invertible Motion Compensated 3-D Subband/Wavelet Filter Bank

Shih-Ta Hsiang, and John W. Woods

Center for Image Processing Research and
Electrical, Computer, and Systems Engineering Department
Rensselaer Polytechnic Institute, Troy, NY 12180-3590

1. ABSTRACT

In this work, we present a new embedded video coding system with *motion compensation*. The invertible motion compensated 3-D subband filter bank is utilized for video analysis. The efficient embedded image coding scheme EZBC is extended to 3-D (spatiotemporal) coding of video subbands. The resulting bitstream is rate scalable and embedded. Simulation results show the proposed algorithm outperforms MPEG-2 by 0.8 - 3.3 dB. It also demonstrates a robust PSNR performance over the rate-scalable coders LZC and 3D-SPIHT for tested image sequences.

2. INTRODUCTION

SNR scalability has become an attractive feature for multimedia applications. With scalability in coding rate, a single bitstream can be forwarded/decoded at various rates to meet different bandwidth requirements in a networking environment. Several successful embedded image coding algorithms have been proposed in the literature [1-4]. These coding schemes not only demonstrate excellent compression performance, they also possess the desirable property of *embeddedness*. By this term we mean that one bitstream generates all the different code rates by simple truncation at any point. Therefore, precise rate control is made possible and the bitstream can be decoded at a virtually infinite number of quality levels.

Due to its recursive coding structure, the conventional hybrid video coder is not particularly suited for SNR scalable coding. With temporal DPCM, the previous reconstructed frame is utilized to predict the current frame. Error propagation, or "prediction drift," can occur at the receiver when the bitstream is decoded at a lower rate. To avoid this problem, multiple prediction loops are generally employed for SNR scalable coding. Otherwise, some drift correction schemes have to be applied [6]. In either case, compression efficiency may be severely degraded. Hence, no more than three SNR layers exist in practice. Diverse bandwidth constraints in a heterogeneous network environment cannot be satisfied in this way.

On the other hand, in addition to its high energy compaction, the 3-D subband transform provides a non-recursive coding structure. It is thus particularly suited for rate scalable coding applications. Several embedded image coding algorithms have been extended using the 3-D subband approach [3, 9, 10]. Unlike the conventional hybrid coder, SNR scalability in these coding systems is

achieved (almost) without loss of coding efficiency, thanks to the embeddedness property of their bitstreams.

Rate scalable 3-D subband video coding was pioneered by Taubman and Zakhor in their layered zero coding (LZC) [3]. With effective exploitation of strong correlation among subband coefficients, LZC reported one of the best PSNR performances for image compression. For video coding, they considered the effects of a camera pan (global motion vector) and pre-distorted the images before temporal analysis filtering. Because only constant displacement motion compensation was adopted, temporal redundancy is not effectively removed when significant trackable local motion exists in the video.

The well-known embedded image coder SPIHT [2] was extended to 3-D or spatiotemporal coding in [9]. This algorithm attempted to exploit the energy clustering property of 3-D subband/wavelet coefficients. The 3-D hierarchical tree structure was utilized for efficient representation of "insignificant coefficients." The low complexity feature of SPIHT was still retained. The coding bitstream was bit-rate scalable and fully embedded. Nevertheless, temporal filtering without motion compensation tends to produce blurred (or double) still images when large motion exists in the input video. If only temporal low frequency subbands are displayed for a low frame-rate rendition, the resulting blurred reconstructed video is visually unacceptable even without any coding distortion. Moreover, the compression efficiency decreases with more energy present in the high temporal subbands.

In Tri-Zerotree [10], Tham *et al.* extended EZW to three dimensions with motion compensation for very low bit rate coding. However, because unconnected or multiply connected pixels in the temporal filtering phase were ignored, their temporal analysis/synthesis system was not perfectly reconstructible and is thus not suitable for high-quality video coding applications.

In this work, we present a new spatiotemporal subband embedded video coding system, referred to IMC-3DEZBC. The MC-3D subband filter bank proposed in [8] is adopted for video decomposition. In this analysis/synthesis system, MC temporal filtering is performed with half-pixel accuracy. The subband filter bank is invertible while retaining high energy compaction, as demonstrated in [8]. For embedded coding of 3-D subband coefficients, we extend our new and efficient image coder EZBC [4] to 3-D coding. The EZBC method combines the advantages of two popular and powerful coding techniques: set partitioning and context modeling. The coded bitstream is fully embedded. The experimental results in [4] showed that EZBC outperformed the zerotree/block coders SPIHT and SPECK [5] and was comparable to the state-of-art JPEG 2000 VM3.1A in compression efficiency.

With zeroblock coding, 3D-EZBC is computationally less complex than LZC for coding bitplanes of subband coefficients. Compared to 3D-SPIHT, IMC-3DEZBC possesses a relatively higher complexity due to the motion estimation. Nevertheless, this operation only needs to be carried out once. The estimated motion vector can be applied to reconstruct video at any admissible quality level.

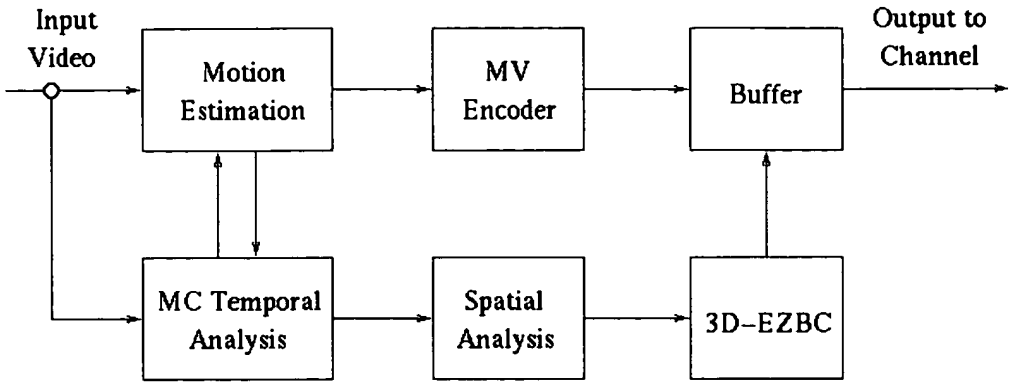


Figure 1 Block diagram of the coding system

This paper is organized as follows: The next section gives an overview of the complete video coding system. Section 4 describes the extension of EZBC to three dimensions for coding 3-D subband/wavelet coefficients. Simulation results are presented in Sec. 5, followed by conclusions.

3. OVERVIEW OF THE CODING SYSTEM

A diagram of the proposed coding system is shown in Fig. 1. The input video is first temporally decomposed by the MC two-channel analysis system. Four stages of the analysis operation are recursively performed on the low frequency subband to generate an octave based five-band decomposition, as shown in Fig. 2. Four-stage spatial analysis follows this temporal stage to complete the 3-D subband decomposition. We adopted Daub. 9/7 [14] analysis/synthesis filters for the spatial filtering. Hierarchical Variable Size Block Matching (HVSBM) [7], which helps reduce the number of unconnected pixels, is utilized for motion estimation. The sizes of our square motion blocks range from 4 x 4 to 64 x 64.

Our coding system divides consecutive frames into groups, similar to the GOP structure in MPEG. Each GOP contains 16 frames – 1 t-LLLL frames, 1 t-LLLH frame, 2 t-LLH frames, 4 t-LH frames, and 8 t-H frames. The 3-D subbband structure of the GOP after transform is illustrated in Fig. 3. Different resolutions and frame rates are available by just decoding the subbands of the corresponding frequencies. The rate control is conducted for each GOP with the bit budget given by

$$Rg = Ng \times R / F \text{ (bits)}$$

where

Ng: the number of frames in a GOP,
R: the coded bit rate in bits/sec,
F: the frame rate of the image sequence in frames/sec.

The 3-D subband coefficients are coded by the 3-D version of EZBC described later. Each GOP is coded as a separate message with context-dependent arithmetic coding. With the feature of embeddedness, the bitstream can be truncated at any point once the desired quality or

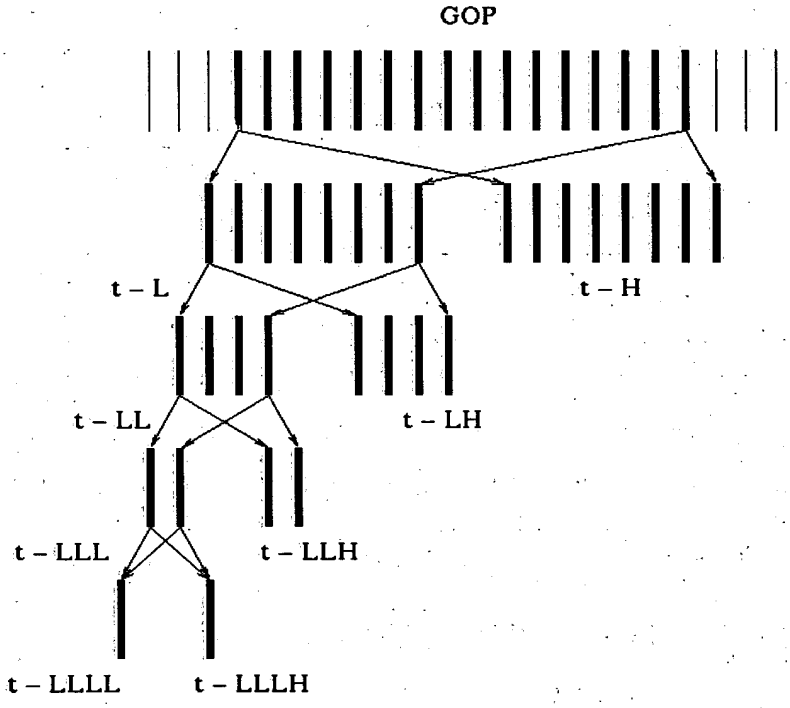
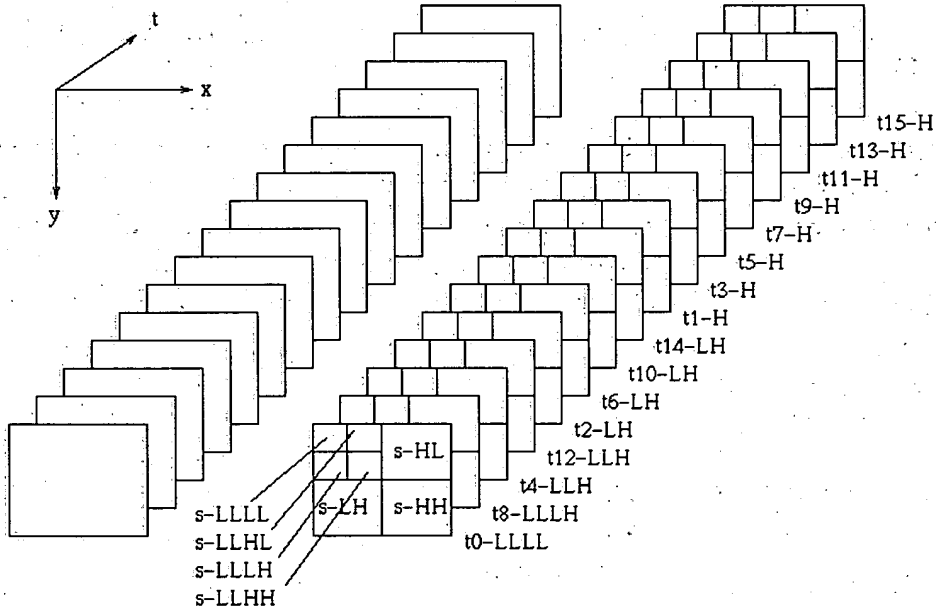


Figure 2 Octave based five-band temporal decomposition



(a) Input Video (b) Decomposition of GOP

Figure 3 The 3-D subband structure in a GOP

the bandwidth constraint is reached. The motion vectors are also coded as side information, which is a"non-scalable"part of the bitstream. For color video coding, the sub-bitstreams for Y, U, and V components are interleaved in each sub-bitplane coding pass, as done in [9].

4. THREE-DIMENSIONAL EZBC

Extension of EZBC to 3-D coding is straightforward, and outlined as the follows:

Definition

- $c(i, j, t)$: integral part of suband/wavelet coefficients at position (i, j, t) after proper scaling.
- $QT_k[l](i, j, t)$: value of quadtree node at position (i, j, t), quadtree level l and subband k.
  $QT_k[0](i, j, t) := |c(i, j, t)|$
  $QT_k[l](i, j, t) := \max(QT_k[l-1](2i, 2j, t), QT_k[l-1](2i, 2j+1, t), QT_k[l-1](2i+1, 2j, t), QT_k[l-1](2i+1, 2j+1, t))$
- $m(i, j, t)$: MSB (most significant bit) of the value of quadtree node (i, j, t).
- D_k: depth of the quadtree of subband k
- $S_n(i, j, t)$: significance of quadtree node (i, j, t) with respect to n, as below
  $S_n(i, j, t) := \begin{cases} 1, & n <= m(i,j, t) \\ 0, & \text{otherwise} \end{cases}$
- $LIN_k[l]$: list of insignificant nodes from quadtree level l of subband k
- LSP_k: list of significant pixels from subband k

Initialization

- $LIN_k[l] = \begin{cases} \{ (0, 0, t) \}, & l = D_k, \text{ any } t \\ \text{empty}, & \text{otherwise.} \end{cases}$
- LSP_k = empty.

Coding_of_LIN(k, l, n)**{**
  for each (i, j, t) in $LIN_k[l]$
    code $S_n(i, j, t)$
    if $(S_n(i, j, t) == 0)$
      (i, j, t) remains in $LIN_k[l]$
    else
      if $(l == 0)$
        code sign bit of $c(i, j, t)$
        add (i, j, t) to LSP_k
      else
        Code_Just_Signif_Node(i, j, t, k, l, n)**}**

Coding_of_LSP(*k, n*){
  for each pixel *(i, j, t)* in LSP_k
    code the *n*-th bit of *c(i, j, t)* }

Code_Just_Signif_Node(*i, j, t, k, l, n*){
  for each node *(x, y, t)* in *{ (2i, 2j, t), (2i, 2j+1, t), (2i+1, 2j, t), (2i+1, 2j+1, t)}*, quadtree level *l*, subband *k*
   code $S_n(x, y, t)$
   if*(Sn(x, y, t) == 0)*
     add *(x, y, t)* to $LIN_k[l-1]$
   else
     if*(l == 1)*
       code the sign bit of *c(x, y, t)*
     else
       Code_Just_Signif_Node*(x, y, t, k, l-1, n)* }

The 3D-EZBC algorithm begins with establishment of the quadtree representations of individual subbands for each frame. The value of a quadtree node is just equal to the maximum magnitude of the subband/wavelet coefficients in this node's corresponding block region. In contrast with conventional pixel-wise bitplane coding algorithms such as [3, 12], EZBC also needs to deal with bitplanes of nodes at individual quad-tree levels. Nevertheless, instead of all subband pixels, only nodes in lists LIN and LSP, a much smaller number, have to be processed in each bitplane coding pass.

To code the significance of the quadtree nodes, we include 8 neighbor nodes of the same quadtree level in the spatial context, as illustrated in Fig. 4. This spatial context has been widely adopted for coding of the significance of subband/wavelet coefficients, e.g. EBCOT [12]. However, the information across scale (context) is given by the node of the spatial parent subband at the next lower quadtree level (the next higher resolution level), as shown in the Fig. 4. This choice is based on the fact that at the same quadtree level the dimension of the parent subband is halved as a result of sub-sampling at the transform stage. The model selection of the arithmetic coding is just based on a 9-bit string with each bit indicating the significance status of nodes in this context. To lower model cost, instead of including all context states, we adopted a similar method to EBCOT for context reduction. The sign coding scheme of EBCOT is also employed in our algorithm.

5. SIMULATION RESULTS

The proposed embedded video coding algorithm has been implemented in software. The test video clips *Mobile Calendar*, *Flower Garden* and *Table Tennis* in SIF resolution (progressively scanned, 352 x 240, 30 fps) were used in our experiments. Each test sequence contains 96 frames.

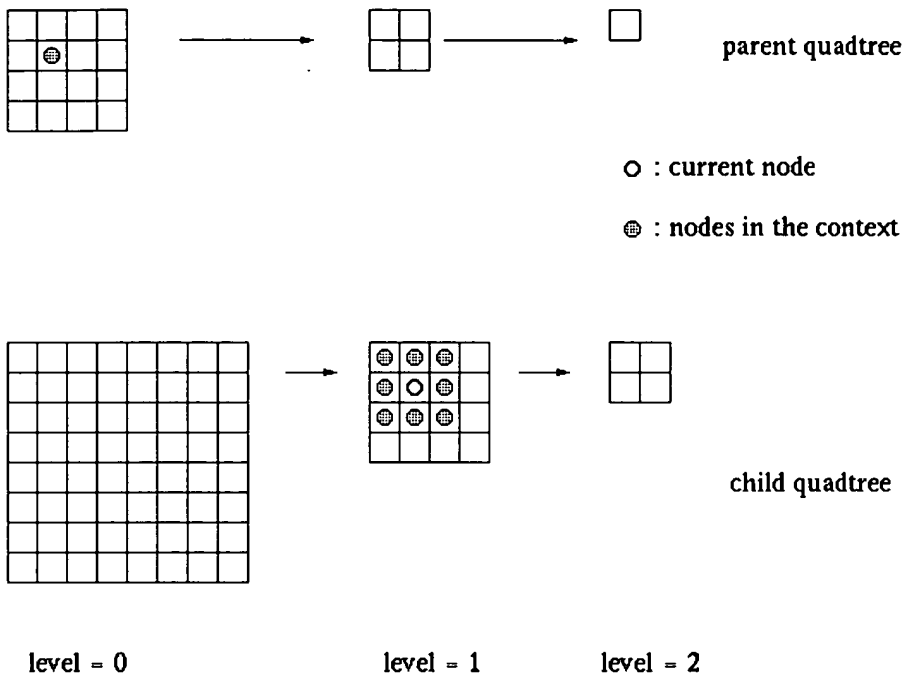


Figure 4 Illustration of the context models of a quadtree node.

In Tables 1 and 2, we compare average PSNR performance for IMC-3DEZBC, MPEG-2 [11], and our earlier 3-D subband coder IMC-3DSBC [8], which uses 3D-FSSQ [7] for coding the subband coefficients. Although PSNR improvement over IMC-3DSBC is not substantial, IMC-EZBC possesses the attractive feature of SNR scalability and the bitstream is fully embedded.

Comparison was also made with two rate scalable video coders 3D-SPIHT and LZC. For LZC coding, we setup the structure file based on [3]; that is, using the nine-tap filters of Adelson *et al.* [13] for spatial filtering, the subband decomposition structure suggested in Fig. 3, pp 575 of [3], and the number of frames for each transmission block $F = 32$. The coder (transmitter) options –pan –minhead were turned on for best compression efficiency.

Table 3 summarizes the average PSNRs for coding gray-level test sequences at bit rates 0.6, 1.2, and 2.4 Mbps. It is clearly shown that the proposed IMC-3DEZBC demonstrates excellent and robust compression performance. Although performing very well on Table Tennis and Mobile Calendar with global motion compensation, LZC performs relatively poorly on Flower Garden. This is owing to the fact that the simple camera pan model is not capable of describing the complicated motion in the Flower Garden sequence. Without motion compensation, 3D-SPIHT shows relatively worse performance on Mobile Calendar and Flower Garden, both clips with large motion.

The average PSNR performance for coding only the luminance component of Flower Garden at various rates is given in Fig. 5. We note that the bitstreams produced by these three coders are all

embedded. Therefore, the coding results shown in Table 3 and Figure 5 for these coders are actually respectively generated by repeatedly decoding a single bitstream at different rates.

6. CONCLUSION

In this research, we presented a 3-D subband/wavelet embedded video coding system. The efficient image coding algorithm EZBC was extended to three dimensions for coding the spatiotemporal subbands. Some attractive features of EZBC such as high compression efficiency, low complexity, and SNR scalability are retained. High energy compaction of transform coefficients, achieved by the invertible MC 3-D subband filter bank, furthers the coding performance. The simulation results show that the proposed coding system outperforms the standard MPEG-2 coder in compression efficiency. It also demonstrates more robust PSNR performance than the other two well-known rate scalable coders LZC and 3D-SPIHT.

References

[1] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3445-3462, Dec. 1993.

[2] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchal trees," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp 243-250, June 1996.

[3] D. Taubman and A. Zakhor, "Multirate 3-D Subband Coding of Video," *IEEE Trans Image Processing*, vol. 3, no. 5, pp. 572-588. Sept. 1994.

[4] S.-T. Hsiang and J. W. Woods, "Embedded image coding using zeroblocks of subband / wavelet coefficients and context modeling," accepted , *ISCAS-2000*, May 2000. http://www.cipr.rpi.edu/publications/pulications.html

[5] A. Islam and W. A. Pearlman,"An embedded and efficient low-complexity hierarchical image coder," VCIP'99, Vol. 3653, pp 294-305, San Jose, CA, 1999.

[6] R. Mathew and Arnold, "Layered coding using bitstream decomposition with drift correction," *IEEE Trans on Circuits and Systems for Video Technology*, vol. 7, no. 6, pp. 882-891, Dec. 1997.

[7] S. Choi and J. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans Image Processing*, vol. 8, no. 2, pp. 155-167, Feb. 1999.

[8] S.-T. Hsiang and J. W. Woods, "Invertible three-dimensional analysis/synthesis system for video coding with half-pixel-accurate motion compensation," *Proc. SPIE Conf. on Visual Communications and Image Processing*, San Jose, CA, 1999.

[9] B. J. Kim and W. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," *Proc. DCC'97, IEEE Data Compression Conference*, pp. 251-260, snowbird, UT Mar. 1997.

[10] J. Y. Tham, S. Ranganathm, and A. A. Kassim, "Highly Scalable Wavelet-Based Video Coding for Very Low Bit-rate Environment," Very Low Bit-Rate Video Coding II, *IEEE Journal on Selected Areas in Communications*, Jan.1998.

[11] MPEG Software simulation group, MPEG-2 TM5 encoder v.1.1, available http://www.mpeg.org/index.html/MSSG/#source.

[12] D. Taubman, "EBCOT (Embedded Block Coding with Optimized Truncation): A complete reference," ISO/IEC JTC1/SC29/WG1 N988, Sept. 1998

[13] E. H. Adelson, E. Simoncelli, and R. Hingorani, "Orthogonal pyramid transform for image coding," in *Proc. SPIE* , Cambridge, MA, Oct. 1987, pp50-58.

[14] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205-220, Apr. 1992.

Image sequence	Coder	Y(dB)	U(dB)	V(dB)
Mobile Calendar	MPEG	22.24	26.54	26.99
	IMC-3DSBC	25.24	29.96	29.62
	IMC-3DEZBC	25.49	30.17	30.72
Flower Garden	MPEG	23.77	30.55	27.65
	IMC-3DSBC	24.88	29.44	29.92
	IMC-3DEZBC	24.95	32.10	29.83

Table 1: PSNR evaluation at 730 Kbps bit rate.

Image sequence	Coder	Y(dB)	U(dB)	V(dB)
Mobile Calendar	MPEG	28.06	31.52	31.79
	IMC-3DSBC	30.64	34.97	34.62
	IMC-3DEZBC	30.88	35.23	35.64
Flower Garden	MPEG	30.32	33.95	32.54
	IMC-3DSBC	30.80	34.80	35.78
	IMC-3DEZBC	31.08	36.53	35.37

Table 2: PSNR evaluation at 2.4 Mbps bit rate.

Image sequence	Coder	0.6 (Mbps)	1.2 (Mbps)	2.4 (Mbps)
Table Tennis	LZC	31.54	34.80	38.66
	3D-SPIHT	30.83	34.23	38.14
	IMC-3DEZBC	31.21	34.72	38.70
Mobile Calendar	LZC	24.04	27.49	32.27
	3D-SPIHT	21.42	23.97	28.02
	IMC-3DEZBC	25.22	28.15	32.04
Flower Garden	LZC	22.22	24.88	29.22
	3D-SPIHT	20.88	23.45	27.36
	IMC-3DEZBC	24.45	27.74	32.14

Table 3: PSNR evaluation at various bit rates for gray-level test sequences.

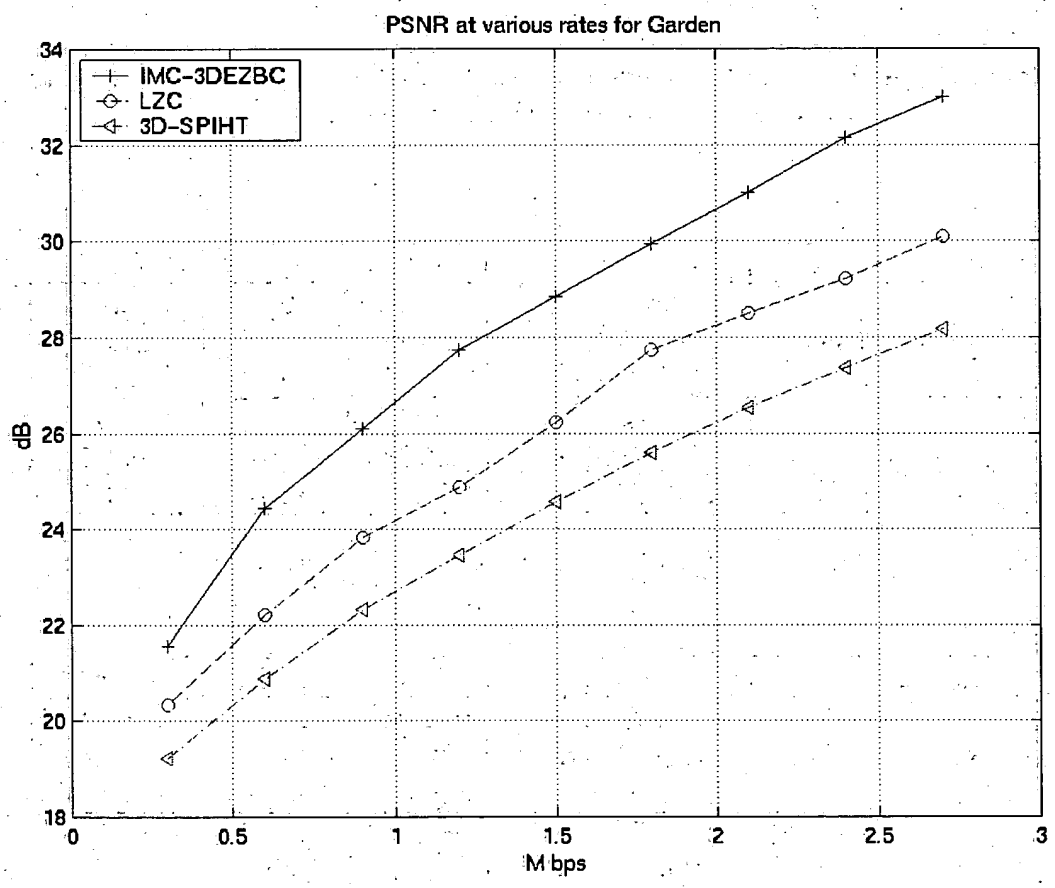


Figure 5: PSNR performance comparison for gray-level Flower Garden.